wwPDB EM Validation Summary Report ⓘ

May 7, 2024 – 02:55 pm BST

PDB ID	:	8S54
EMDB ID	:	EMD-19720
Title	:	RNA polymerase II early elongation complex bound to TFIIE and TFIIF - state b (composite structure)
Authors	:	Zhan, Y.; Grabber, F.; Oberbeckmann, E.; Dienemann, C.; Cramer, P.
Deposited on	:	2024-02-22
Resolution	:	3.40 Å(reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

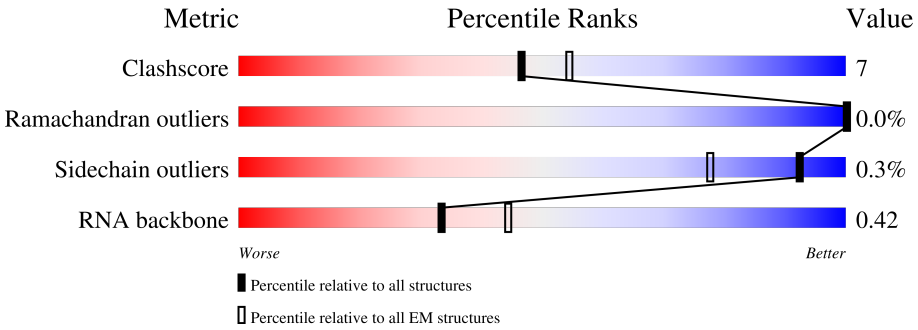
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.40 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



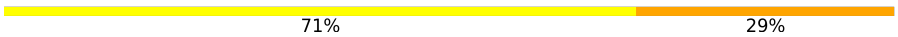
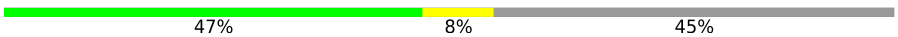
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	1984	59% 12% 29%
2	B	1300	72% 15% 13%
3	C	275	78% 15% 7%
4	D	184	63% 7% 30%
5	E	210	86% 14%
6	F	127	48% 17% 35%
7	G	172	90% 10% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	H	150	
9	I	125	
10	J	67	
11	K	117	
12	L	58	
13	N	139	
14	P	14	
15	Q	517	
16	R	249	
17	T	139	
18	W	439	

2 Entry composition (i)

There are 20 unique types of molecules in this entry. The entry contains 36745 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA-directed RNA polymerase subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	1417	11234	7070	2010	2082	72	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	1130	9030	5714	1587	1665	64	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	C	257	2059	1294	351	408	6	0	0

- Molecule 4 is a protein called RNA polymerase II subunit D.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	D	128	1050	656	178	212	4	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II subunit E.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	E	209	1720	1089	300	323	8	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	F	82	657	418	113	121	5	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	171	Total	C	N	O	S	0	0
			1351	875	219	249	8		

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	148	Total	C	N	O	S	0	0
			1186	750	194	237	5		

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	114	Total	C	N	O	S	0	0
			927	571	166	179	11		

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	67	Total	C	N	O	S	0	0
			533	345	90	92	6		

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB11-a.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	114	Total	C	N	O	S	0	0
			916	591	151	172	2		

- Molecule 12 is a protein called RNA polymerase II subunit K.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	44	Total	C	N	O	S	0	0
			372	231	72	63	6		

- Molecule 13 is a DNA chain called Non-template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	N	35	Total	C	N	O	P	0	0
			712	339	132	207	34		

- Molecule 14 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	P	14	Total	C	N	O	P	0	0
			296	133	56	93	14		

- Molecule 15 is a protein called General transcription factor IIF subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	Q	138	Total	C	N	O	S	0	0
			1138	719	208	208	3		

- Molecule 16 is a protein called General transcription factor IIF subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	R	137	Total	C	N	O	S	0	0
			1070	666	193	209	2		

- Molecule 17 is a DNA chain called Template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	T	46	Total	C	N	O	P	0	0
			949	449	175	279	46		

- Molecule 18 is a protein called General transcription factor IIE subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
18	W	187	Total	C	N	O	S	0	0
			1535	964	275	285	11		

- Molecule 19 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		AltConf
19	A	2	Total	Zn	0
			2	2	
19	B	1	Total	Zn	0
			1	1	
19	C	1	Total	Zn	0
			1	1	
19	I	2	Total	Zn	0
			2	2	
19	J	1	Total	Zn	0
			1	1	
19	L	1	Total	Zn	0
			1	1	

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	AltConf
19	W	1	Total Zn 1 1	0

- Molecule 20 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

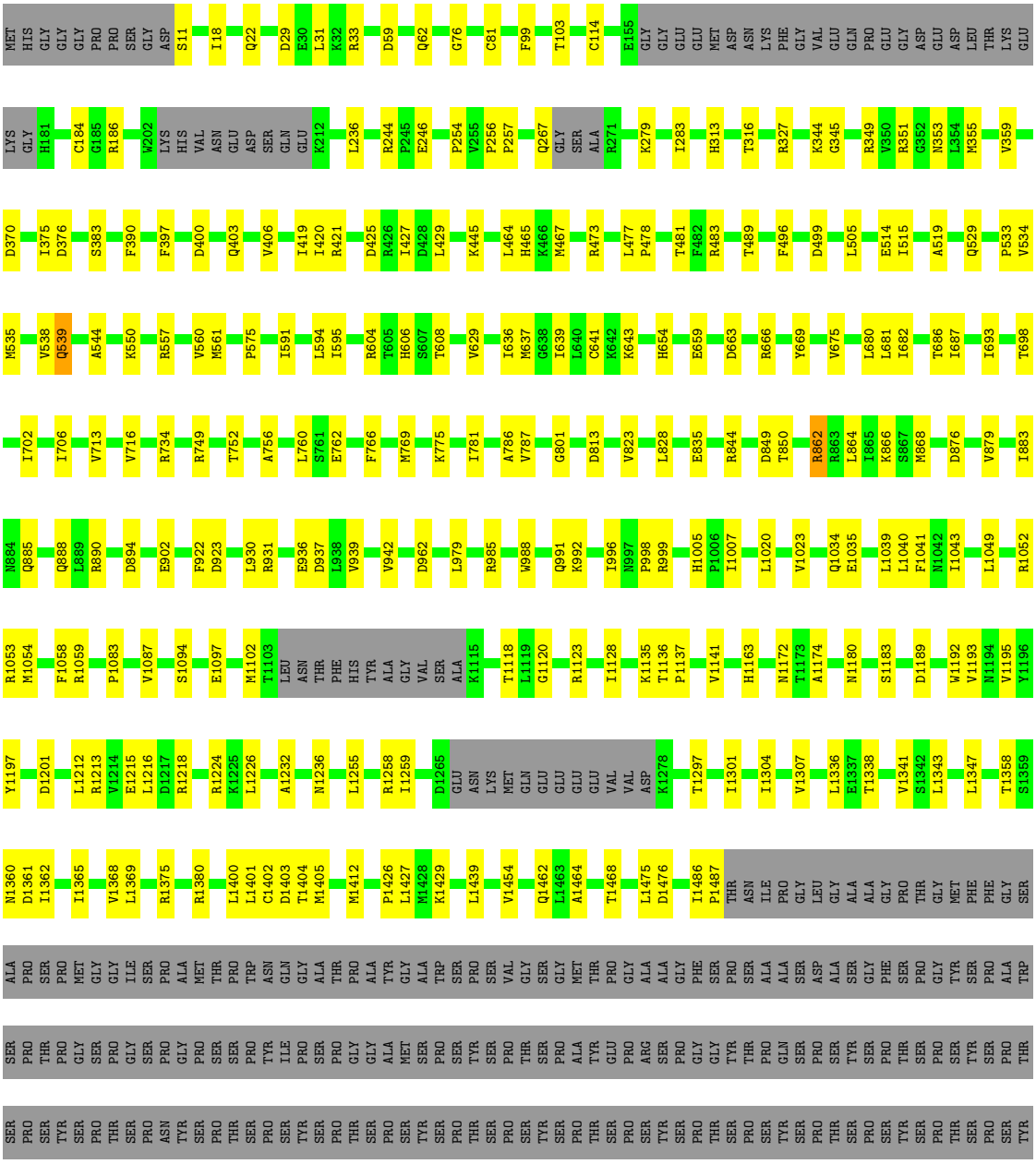
Mol	Chain	Residues	Atoms	AltConf
20	P	1	Total Mg 1 1	0

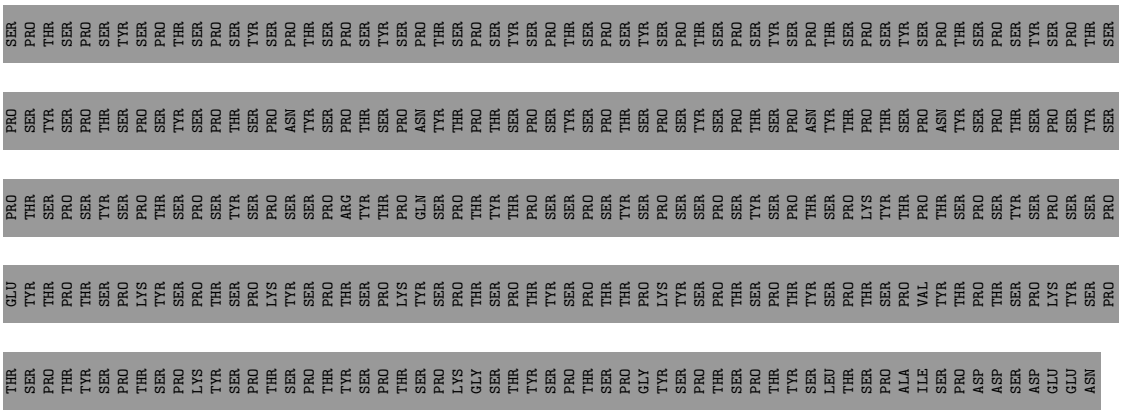
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: DNA-directed RNA polymerase subunit

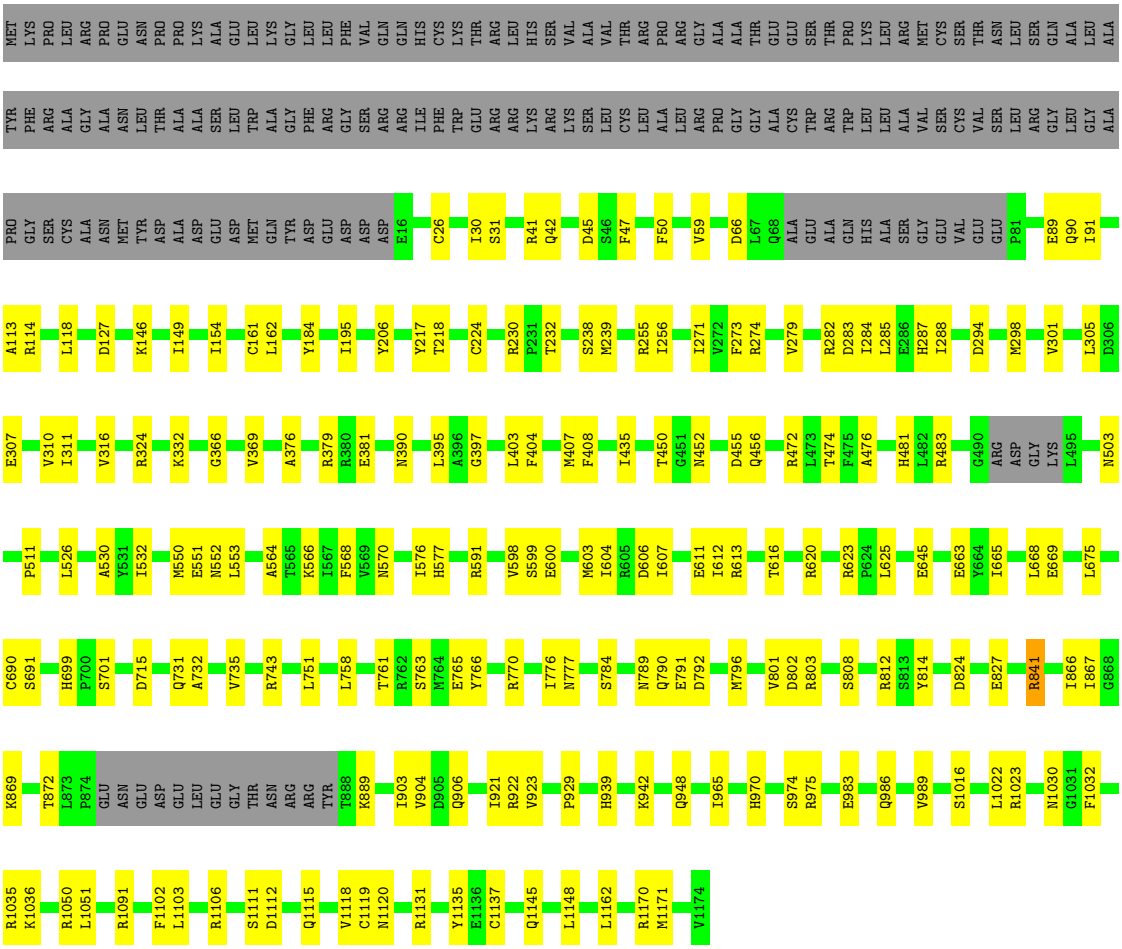
Chain A:





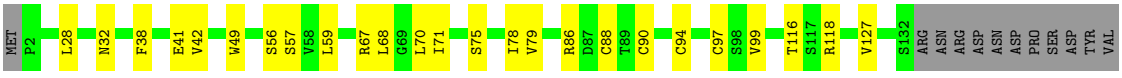
• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain B:

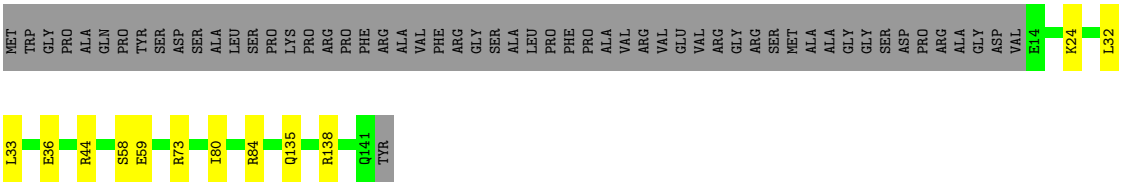


• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

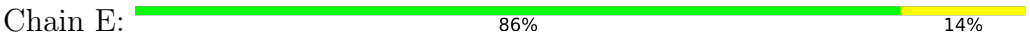
Chain C:



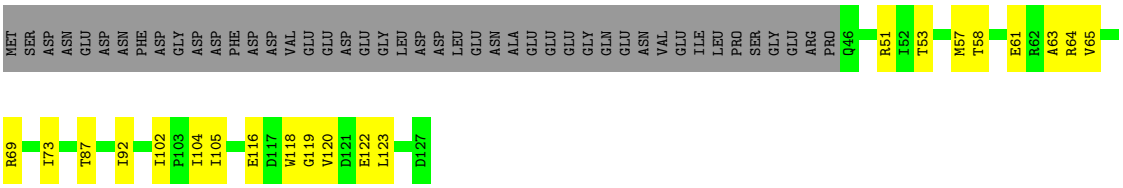
• Molecule 4: RNA polymerase II subunit D



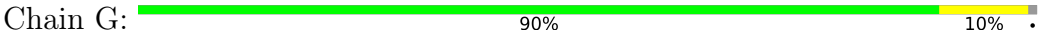
• Molecule 5: DNA-directed RNA polymerase II subunit E

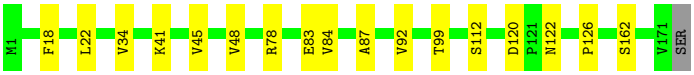


• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC2

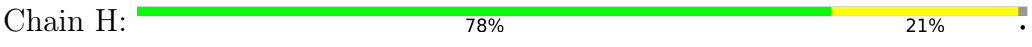


• Molecule 7: DNA-directed RNA polymerase II subunit RPB7

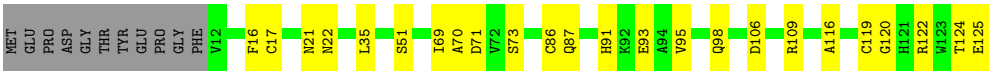




• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC3

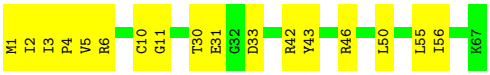


• Molecule 9: DNA-directed RNA polymerase II subunit RPB9



• Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:

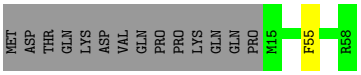


• Molecule 11: DNA-directed RNA polymerase II subunit RPB11-a

Chain K:

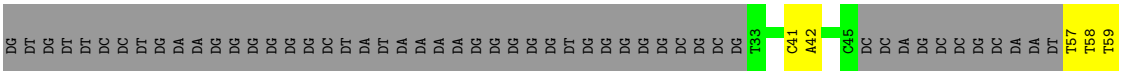
• Molecule 12: RNA polymerase II subunit K

Chain L:



• Molecule 13: Non-template DNA

Chain N:

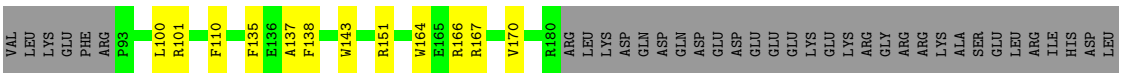


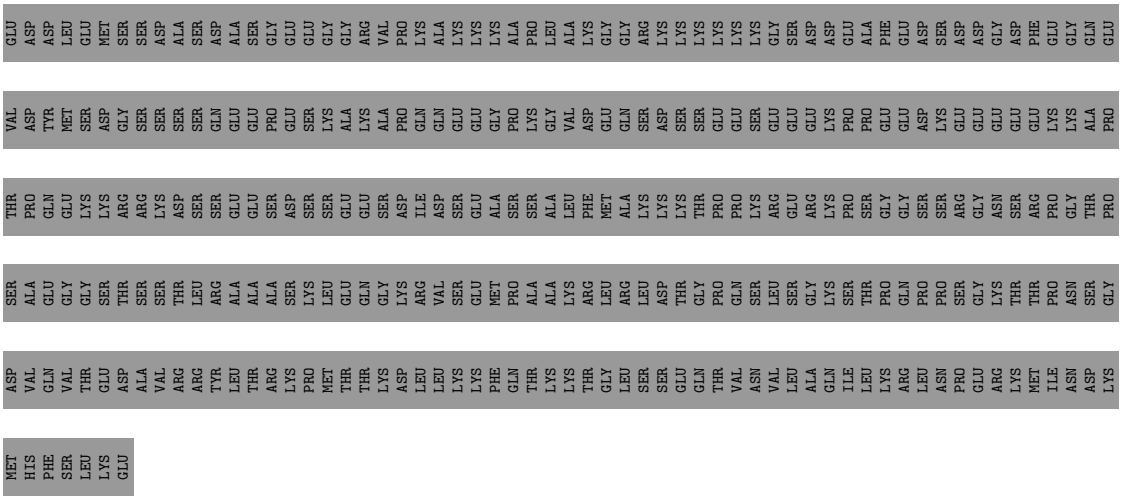
• Molecule 14: RNA

Chain P:

• Molecule 15: General transcription factor IIF subunit 1

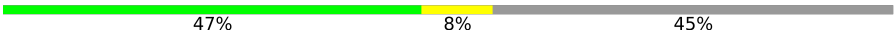
Chain Q:

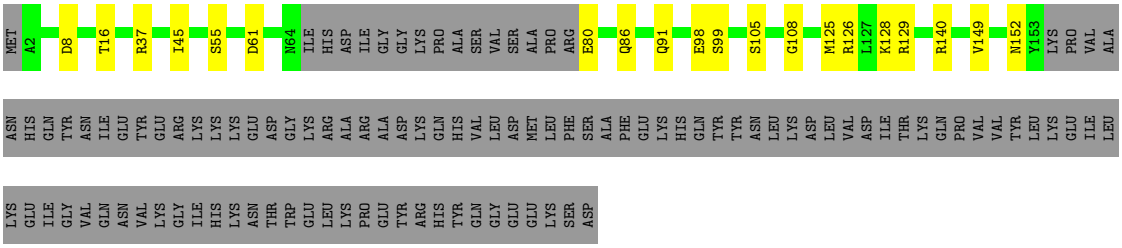




• Molecule 16: General transcription factor IIF subunit 2

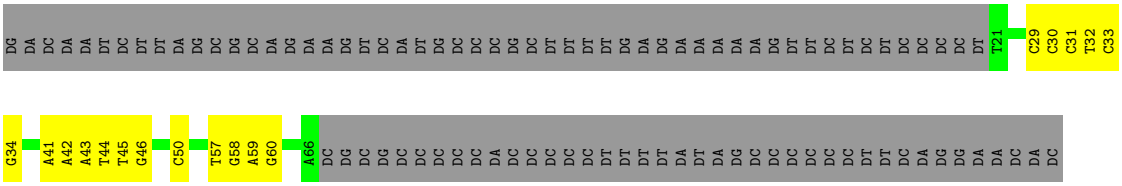
Chain R:





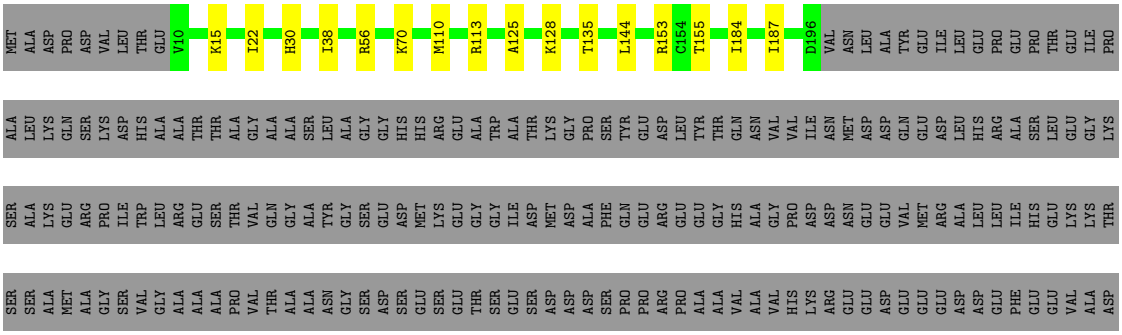
• Molecule 17: Template DNA

Chain T:



• Molecule 18: General transcription factor IIE subunit 1

Chain W:



ASP
PRO
ILE
VAL
MET
VAL
ALA
GLY
ARG
PRO
PHE
SER
TYR
SER
GLU
VAL
SER
GLN
ARG
PRO
GLU
LEU
VAL
ALA
GLN
MET
THR
PRO
GLU
GLU
LYS
GLU
ALA
TYR
ILE
ALA
MET
GLY
GLN
ARG
MET
PHE
GLU
ASP
LEU
PHE
GLU

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	50000	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	40	Depositor
Minimum defocus (nm)	700	Depositor
Maximum defocus (nm)	1700	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	A	0.30	0/11439	0.63	0/15440
2	B	0.29	0/9210	0.62	0/12431
3	C	0.27	0/2102	0.57	0/2857
4	D	0.26	0/1064	0.52	0/1428
5	E	0.29	0/1751	0.63	0/2366
6	F	0.29	0/667	0.68	0/901
7	G	0.28	0/1382	0.57	0/1874
8	H	0.30	0/1207	0.63	0/1628
9	I	0.30	0/948	0.61	0/1284
10	J	0.32	0/542	0.64	0/730
11	K	0.29	0/935	0.58	0/1266
12	L	0.27	0/377	0.68	0/500
13	N	0.96	0/797	1.30	0/1225
14	P	0.75	0/330	1.18	0/511
15	Q	0.28	0/1167	0.61	0/1576
16	R	0.25	0/1083	0.54	0/1459
17	T	0.89	0/1064	1.27	0/1642
18	W	0.26	0/1560	0.57	0/2097
All	All	0.36	0/37625	0.68	0/51215

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
2	B	0	3
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	862	ARG	Sidechain
2	B	472	ARG	Sidechain
2	B	812	ARG	Sidechain
2	B	841	ARG	Sidechain

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	11234	0	11365	167	0
2	B	9030	0	9073	133	0
3	C	2059	0	2007	27	0
4	D	1050	0	1033	7	0
5	E	1720	0	1737	31	0
6	F	657	0	684	15	0
7	G	1351	0	1358	10	0
8	H	1186	0	1147	21	0
9	I	927	0	859	15	0
10	J	533	0	553	12	0
11	K	916	0	939	19	0
12	L	372	0	380	1	0
13	N	712	0	395	5	0
14	P	296	0	154	19	0
15	Q	1138	0	1103	18	0
16	R	1070	0	1086	16	0
17	T	949	0	518	14	0
18	W	1535	0	1540	8	0
19	A	2	0	0	0	0
19	B	1	0	0	0	0
19	C	1	0	0	0	0
19	I	2	0	0	0	0
19	J	1	0	0	0	0
19	L	1	0	0	0	0
19	W	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
20	P	1	0	0	0	0
All	All	36745	0	35931	472	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 7.

The worst 5 of 472 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:E:29:THR:CG2	5:E:32:GLU:CG	2.24	1.15
5:E:29:THR:CG2	5:E:32:GLU:HG3	1.81	1.11
5:E:29:THR:HG22	5:E:32:GLU:CD	1.74	1.07
3:C:90:CYS:SG	3:C:94:CYS:HB3	1.97	1.03
1:A:557:ARG:O	1:A:561:MET:HB2	1.61	1.00

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	1405/1984 (71%)	1322 (94%)	82 (6%)	1 (0%)	51 82
2	B	1122/1300 (86%)	1064 (95%)	58 (5%)	0	100 100
3	C	253/275 (92%)	242 (96%)	11 (4%)	0	100 100
4	D	126/184 (68%)	122 (97%)	4 (3%)	0	100 100
5	E	207/210 (99%)	197 (95%)	10 (5%)	0	100 100
6	F	80/127 (63%)	77 (96%)	3 (4%)	0	100 100
7	G	169/172 (98%)	164 (97%)	5 (3%)	0	100 100
8	H	146/150 (97%)	133 (91%)	13 (9%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	I	112/125 (90%)	98 (88%)	14 (12%)	0	100	100
10	J	65/67 (97%)	61 (94%)	4 (6%)	0	100	100
11	K	112/117 (96%)	103 (92%)	9 (8%)	0	100	100
12	L	42/58 (72%)	36 (86%)	6 (14%)	0	100	100
15	Q	134/517 (26%)	127 (95%)	7 (5%)	0	100	100
16	R	133/249 (53%)	131 (98%)	2 (2%)	0	100	100
18	W	185/439 (42%)	184 (100%)	1 (0%)	0	100	100
All	All	4291/5974 (72%)	4061 (95%)	229 (5%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	539	GLN

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	1249/1763 (71%)	1247 (100%)	2 (0%)	93	98
2	B	990/1127 (88%)	987 (100%)	3 (0%)	92	97
3	C	234/252 (93%)	234 (100%)	0	100	100
4	D	118/160 (74%)	116 (98%)	2 (2%)	60	80
5	E	191/192 (100%)	190 (100%)	1 (0%)	88	94
6	F	71/111 (64%)	71 (100%)	0	100	100
7	G	152/153 (99%)	152 (100%)	0	100	100
8	H	129/131 (98%)	129 (100%)	0	100	100
9	I	103/112 (92%)	103 (100%)	0	100	100
10	J	56/56 (100%)	56 (100%)	0	100	100
11	K	104/106 (98%)	104 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
12	L	41/55 (74%)	41 (100%)	0	100 100
15	Q	121/448 (27%)	120 (99%)	1 (1%)	81 91
16	R	118/218 (54%)	118 (100%)	0	100 100
18	W	169/373 (45%)	167 (99%)	2 (1%)	71 85
All	All	3846/5257 (73%)	3835 (100%)	11 (0%)	92 97

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	47	LYS
15	Q	151	ARG
18	W	153	ARG
18	W	56	ARG
2	B	1131	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	93	HIS
15	Q	32	ASN
8	H	46	GLN
9	I	91	HIS
1	A	1163	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
14	P	13/14 (92%)	3 (23%)	2 (15%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
14	P	2	C
14	P	3	A
14	P	7	A

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
14	P	2	C
14	P	6	C

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 10 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Map visualisation

This section contains visualisations of the EMDB entry EMD-19720. These allow visual inspection of the internal detail of the map and identification of artifacts.

Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

This section was not generated.

6.2 Central slices

This section was not generated.

6.3 Largest variance slices

This section was not generated.

6.4 Orthogonal standard-deviation projections (False-color)

This section was not generated.

6.5 Orthogonal surface views

This section was not generated.

6.6 Mask visualisation

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

This section contains the results of statistical analysis of the map.

7.1 Map-value distribution

This section was not generated.

7.2 Volume estimate versus contour level

This section was not generated.

7.3 Rotationally averaged power spectrum

This section was not generated. The rotationally averaged power spectrum had issues being displayed.

8 Fourier-Shell correlation

This section was not generated. No FSC curve or half-maps provided.

9 Map-model fit

This section was not generated.